



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Transforming F0 Contours

Citation for published version:

Gillett, B & King, S 2003, Transforming F0 Contours. in *Eurospeech 2003 - Interspeech 2003: 8th European Conference on Speech Communication and Technology*. International Speech Communication Association, pp. 101-104.

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Eurospeech 2003 - Interspeech 2003

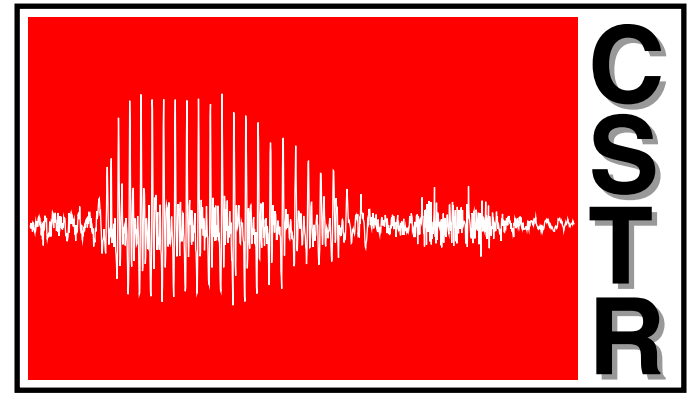
General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





Transforming F0 Contours



Ben Gillett, now with Camel Audio (www.camelaudio.com)

Simon King, Centre for Speech Technology Research, University of Edinburgh, UK

ben@camelaudio.com, Simon.King@ed.ac.uk

Introduction

Goal

To transform the F0 contour of some speech from a source speaker, such that listeners believe it to have been uttered by some target speaker

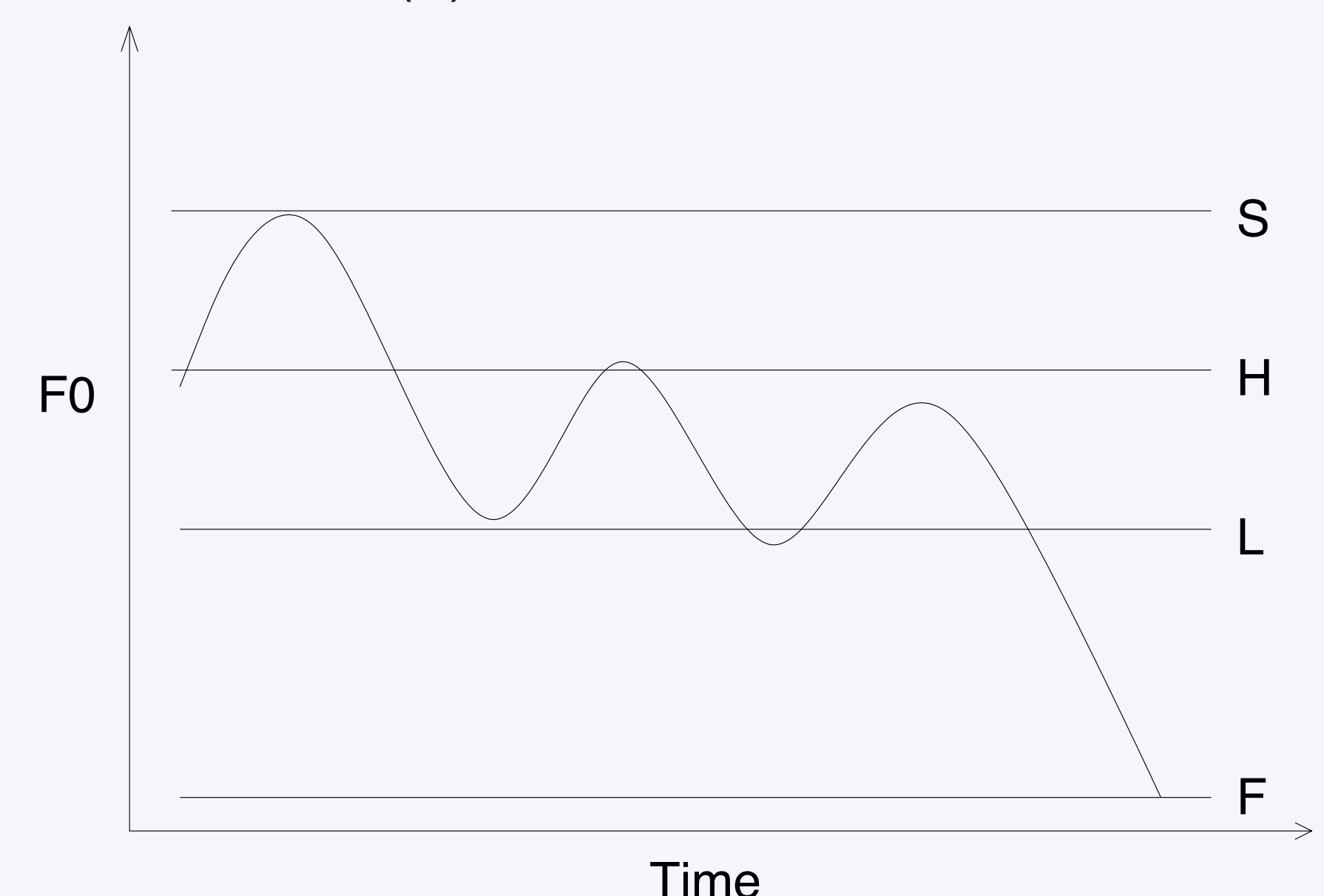
Applications

- voice transformation
 - also need a method for transforming voice quality - see poster by King & Gillett in session PWeBe
- speech synthesis
 - as a way of adapting existing intonation models (trained on one speaker) to a new speaker, without having to annotate much more data

Parameter set

After Patterson

- sentence-initial high (S)
- non-initial accent peaks (H)
- post-accent valleys (L)
- sentence-final low (F)



Mapping

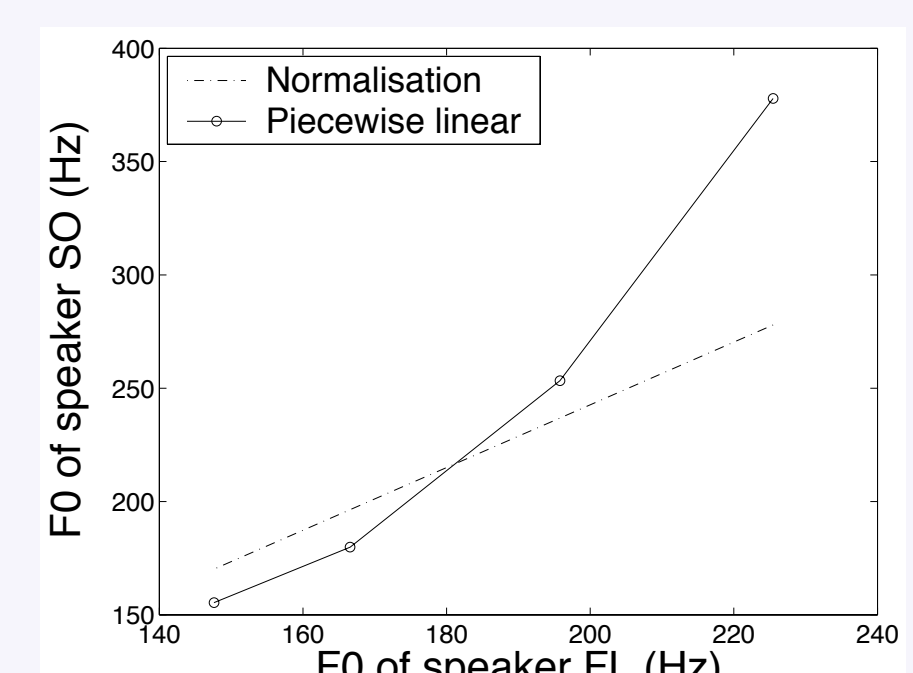
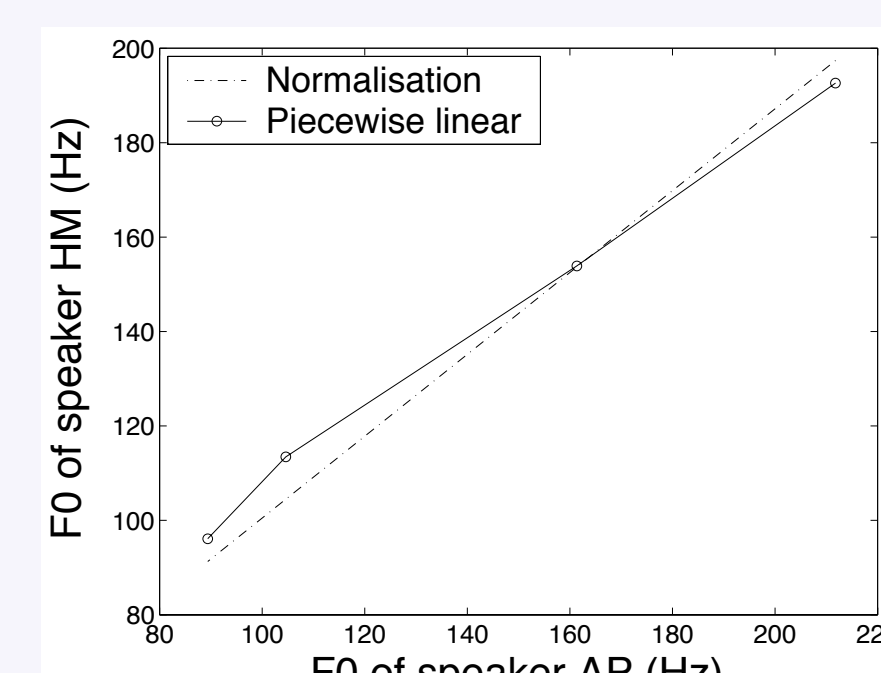
Standard method

Usual method of normalising F0 is to use this mapping

$$M_N(x) = ((x - \mu_{src}) / \sigma_{src}) * \sigma_{targ} + \mu_{targ}$$

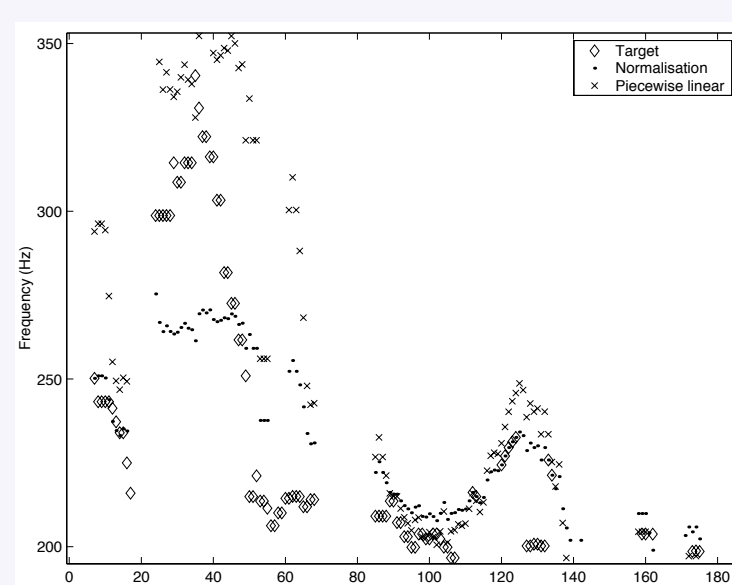
New method

A nonlinear mapping, M_{PL} , composed of piecewise linear sections between F, L, H and S

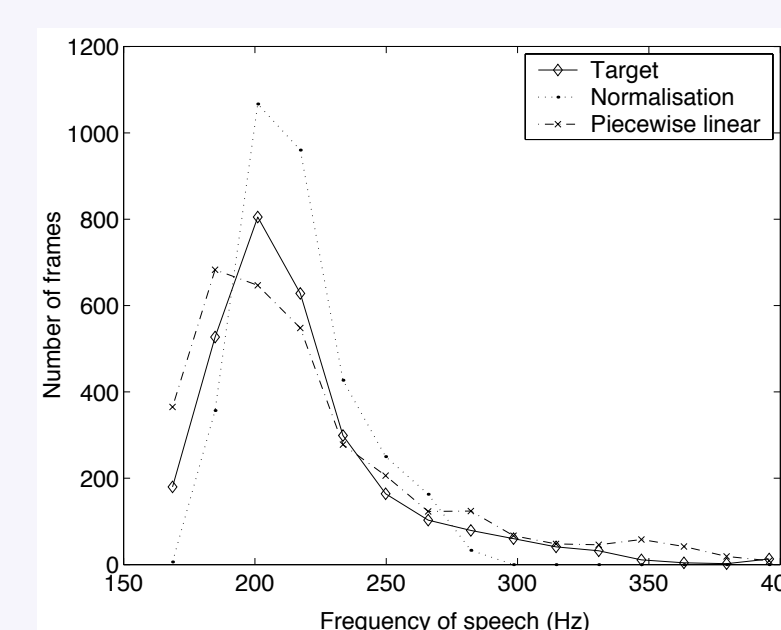


Examples

Transformed contours



Histograms



Audio examples

Evaluation

Perceptual experiment

25 subjects presented with speech from target speaker and speech with transformed F0 contours (“imitator speech”) in XABX format. Asked to judge which of A or B was most like X. A and B varied between: actual target F0, normalised F0 (standard method), transformed F0 (new method).

Speaker pairs classified as “similar” (S_{same}) or “different” ($S_{different}$).

Results

	Mean (%)	Std. Dev.	α	t
Preference for M_{PL} over M_N for $S_{different}$	67	10	$< 1 \times 10^{-7}$	-8.71
Preference for M_{PL} over M_N for S_{same}	54	8	~ 0.02	-2.49
Preference for target over mapped contours	73	9	$< 1 \times 10^{-11}$	-13.8

What next?

Apply the method to full voice transformation or speech synthesis

See also...

- at this conference:
 - poster by Gillett & King in session PWeBe (voice *quality* transformation)
- www.cstr.ed.ac.uk for latest progress on voice transformation and speech synthesis
- www.camelaudio.com for musical instrument transformation and morphing